Relationship of Forecast Encompassing to Composite Forecasts with Simulations and an Application

Kuo-yuan Liang and Keunkwan Ryu [1]

This paper examines the role of forecast-encompassing principles in model-specification searches through the use of linear composite forecasts. Based on the results of the pairwise forecast-encompassing test, this paper outlines a conceptual framework to provide some useful insights on cross-model evaluations in econometrics and the selection of predictors in composite forecasts. Second, it offers three different ways of performing the encompassing test and compares their finite sample performance through a Monte Carlo simulation study. Test results guide researchers to choose component forecasts and thus to avoid blind pooling in the combining regression.

Keywords: Choice of forecasts, Composite forecasts, Forecast-encompassing test, Monte-Carlo simulation, Non-nested hypotheses

JEL Classification: C52, C53, C12, C15

I. Introduction

In the literature on linear combination of forecasts (*e.g.*, Granger and Ramanathan (1984)) and cross-model comparisons of economic forecasts (*e.g.*, Nelson (1972) and Wallis (1989)), it has become

*President, Polaris Research Institute, 1F, No. 339, SinHu 2nd Road, NeiHu, 114, Taipei, Taiwan, (Tel) +886-2-8791-8899, (Fax) +886-2-2790-6594, (E-mail) kyliang@polaris.com.tw; Associate Professor, School of Economics, Seoul National University, San 56-1, Sillim-dong, Kwanak-gu, Seoul 151-742, Korea, (Tel) +82-2-880-6397, (Fax) +82-2-886-4231, (E-mail) ryu@snu.ac.kr, respectively. The authors would like to thank anonymous referees for helpful comments, and the corresponding author would like to thank the Korea Research Foundation for financial assistance through BK21 funding.
[**Seoul Journal of Economics** 2003, Vol. 16, No. 3]

standard to estimate the combining weights and to assess the performance of forecasts relative to each other by running the following linear regression:

$$y_t = \beta_0 + \beta_1 f_{1t} + \beta_2 f_{2t} + \cdots + \beta_k f_{kt} + u_t \qquad (1)$$

where subscript t indexes the order of observations ($t = 1,2,\cdots,T$), y_t is the outcome series, f_{it} ($i = 1,2,\cdots,k$) is the i^{th} forecast series, β_i ($i = 1,2,\cdots,k$) are unknown parameters, and u_t is the error term. Despite the similarity between composite forecasts and cross-model evaluation, historically they were developed independently (Cooper and Nelson 1975). It is interesting to note that articles on composite forecasts are primarily published in forecasting, management science, and operations research journals; articles on cross-model evaluation of economic forecasts have mostly appeared in economic journals. Additionally, the role of the regression equation (1) in each paradigm is different.

The method of combining forecasts is based on the fact that alternative forecasts of the same variable are often available and on the belief that each of them most likely contains useful information. Under the error-variance minimizing criterion, Bates and Granger (1969) represents the most influential early contribution to the linear combination of forecasts. Later, Granger and Ramanathan (1984) demonstrates that the method can be interpreted as, and generalized to, the estimation of the regression equation (1). Clemen (1989) writes a thorough survey on this theme.

In the case of evaluating economic forecasts, a popular approach is that of Nelson (1972) and Cooper and Nelson (1975). Their proposal reflects the dissatisfaction with the use of some conventional error measures such as root mean squared percentage errors (RMSPEs %) and mean absolute percentage errors (MAPEs %) to rank economic forecasts. It is now generally agreed that even a relatively inaccurate forecasting model could contain useful information which is not shared by other competing models (Nelson 1972). To find out the forecasting abilities of competing models, the Nelson-Cooper procedure essentially uses the linear regression (1) as a benchmark and examines statistical significance of the estimated coefficients.

Chong and Hendry (1986) proposes the concept of encompassing in forecasting as a basis for cross-model comparisons. The statistical procedure consists of regressing the *out-of-sample* forecasting errors from a particular model on a selected rival forecast series and performing the encompassing test through a significance test of the forecast series on the righthand side. In addition to Chong and Hendry's (1986) procedure, we show below that the forecast-encompassing test between two separate models can also be carried out in two alternative ways and that all three methods are based on the combining regression (1) and its variant.

Economists and statisticians were skeptical about the method of composite forecasts, when the idea was first introduced by the group under Granger at the University of Nottingham about three decades ago (see the discussions of Newbold and Granger (1974)). In view of mainstream econometricians, the primary interest in modelling an economic system is to understand the structural relationship. That is, they would be more interested in developing the "*true*" structural model in a particular system. Therefore, Diebold (1989) pointed out that "there is no role for forecast combination within such a paradigm" (p. 590), because if the "*true*" structural model can be constructed and verified then the "*good*" forecast will follow automatically. However, the principle of combining forecasts has been increasingly accepted in the forecasting profession (*e.g.*, Bunn (1989)). The change reflects a view that the pooling approach is pragmatic (Diebold 1989; and Winkler 1989). Although aggregation of information sets is superior to aggregating forecasts, the aggregation of information sets is either impossible or too costly. Alternatively speaking, although combining forecasts is inferior to aggregating information sets, combining forecasts is easy to implement and thus practical.

As a result, most of the theoretical and empirical works on the pooling approach are aimed at showing and demonstrating the superiority of the composite forecasts over the individual forecasts under some optimality criteria such as minimizing mean squared error (MSE) criterion. In contrast, the Nelson-Cooper and Chong-Hendry procedures use the combining regression (1) and its variant to examine the forecasting ability of a particular model, and then to assess whether the model suffers from misspecification. In other words, from an econometric perspective, the combining regression (1) is merely a tool which is mainly designed to examine the strength and weakness of a particular model. Using (1), econometricians can test whether a specification dominates others or suffers from misspecification. However it is not the major concern to them whether the linear combination of forecasts is helpful to achieve predictive accuracy.

The main purpose of this paper is to investigate the relationship of forecast encompassing to composite forecasts. In addition to the general understanding that the combining regression is a useful tool for model specification searches, it shows that the test results based on the forecast-encompassing principle can offer valuable insights on the choice of forecasts in the combining regression. The focus is on evaluating separate (non-nested) models for their relative forecasting ability and on treating the encompassing principle (Mizon and Richard 1986) as a unifying framework to guide the selection of component forecasts in the combining regression.

It should be noted that Diebold (1989) deals with a similar issue in his discussion of Clemen's (1989) paper. Nevertheless, his emphasis is on the role of the combining regression as a hint for model-specification searches but not the other way around. This paper is more general because it argues that there exists a two-way interaction between forecast-encompassing principles and composite forecasts. More specifically, this paper argues that on the one hand the combining regression is a useful tool for model-specification searches; on the other hand the forecast-encompassing principle offers valuable insights on the choice of forecasts in the combining regression. Once the interaction is established, the complementary role of the composite forecasts and forecast-encompassing principle can be understood more clearly.

The organization of this paper is as follows. Section II summarizes the main features of three forecast-encompassing test procedures and the relationships among them. Section III details the implementation of the Wald tests associated with the forecast -encompassing tests. Section IV compares the finite sample performance of three tests using Monte-Carlo simulations. In section V we compare the proposed forecast encompassing test procedures with Davidson and MacKinnon's J-test using aggregate U.S. consumption data for 1929 through 1989. Section VI studies the implications of the forecast encompassing test results and applies them to the issue of model-specification searches in econometrics and choice of forecasts in composite forecasts. Concluding remarks follow in section VII.

II. Forecast-Encompassing Test

This section outlines Chong and Hendry's (1986) forecast-encompassing test procedure and two other versions of the test. Consider the two linear forecasting models with claimed formulations:

$$H_1 : Y = X_1 \alpha + U_1, \quad U_1 \sim N(0, \sigma_1^2 I_T) \tag{2}$$

$$H_2 : Y = X_2 \pi + U_2, \quad U_2 \sim N(0, \sigma_2^2 I_T) \tag{3}$$

where Y is a $T \times 1$ vector of observations on the variable being forecast; X_1 and X_2 are $T \times d_1$ and $T \times d_2$ full column rank matrices of observations on explanatory variables; α and π are $d_1 \times 1$ and $d_2 \times 1$ vectors of unknown parameters; and U_i $(i=1,2)$ is a $T \times 1$ vector of normally, independently and identically distributed random disturbance terms. The normality assumptions are not essential since the test statistics are asymptotic in nature and some version of the central limit theorem is to work. X_1 and X_2 in general may share some common variables; however, they are not nested within each other. That is, for the two hypotheses H_1 and H_2 to be separate, at least one column in X_1 must be linearly independent of columns in X_2, and vice versa. The issue here is to test whether a maintained model can predict the performance of the other model.

Let f_{1t} and f_{2t} denote two m-periods $(m-1,2,\cdots)$ ahead forecast series based on H_1 and H_2, respectively. Chong and Hendry (1986) considers a simplified version of the combining regression (1):

$$y_t = \beta_1 f_{1t} + \beta_2 f_{2t} + u_{t'} \tag{4}$$

Then the forecast-encompassing test of H_1 as the null is to test the null hypothesis that $\beta_1 - 1$ and $\beta_2 - 0$. This testing procedure shares the spirit of Davidson and MacKinnon's (1981) J-test. The J-test tests the significance of the least squares estimator of β_2 in the following regression:

$$Y = X_1 \tilde{\alpha} + \tilde{\beta}_2 F_2 + \tilde{U},$$

where $F_2 = X_2(X_2'X_2)^{-1}X_2'Y$ is the fitted value of Y based on model H_2. However, there are two clear differences. First, we are computing model-specific forecasts recursively, which makes more sense in the case of forecasting. Second, the test of H_1 as the null hypothesis uses as one of the regressors, its own predicted value f_{1t} instead of H_1-specific regressors X_1.

It is worth noting that there are three alternative ways to implement the test.

First, we can directly test the joint hypothesis that $\beta_1=1$ and $\beta_2=0$ using, for example, a Wald test. If the result is not significant, then we cannot reject the null that model 1 encompasses model 2 in forecast; otherwise, we reject the null. In the latter case, notice that model 2 forecasts are not encompassed in model 1. Accordingly it is logical for researchers to extract information from f_2 as well as from f_1.

Second, we can test whether $\beta_2=0$ in the combining regression of (4). If model 2 forecasts do not provide additional information beyond that already contained in model 1 forecasts, the estimate of β_2 will not be significantly different from 0. Intuitively, if the combined forecast based on (4) has an error variance that is not significantly smaller than that based only on f_1, then f_2 appears to provide no useful information beyond that already contained in f_1. In such a case, we conclude that model 1 encompasses model 2 in forecast. This procedure was used in Cooper and Nelson (1975) as well as in Fair and Shiller (1990).

Third, when $\beta_1=1$ is true, (4) is expressible as:

$$y_t-f_{1t}=\beta_2 f_{2t}+u_t \tag{5}$$

Therefore, given $\beta_1=1$, whether the second forecast is capable of explaining the errors in the first forecast can be tested statistically by inspecting the significance of the coefficient β_2. If β_2 is found to be significantly (insignificantly) different from zero, then the null H_1 is rejected (not rejected) and model 1 forecasts are said to be incapable of encompassing (capable of encompassing) model 2 forecasts. This is the procedure proposed in Chong and Hendry (1986). Empirical studies using this procedure in cross evaluations of macroeconometric models include Fisher and Wallis (1990), and Charemza (1991).

The second and third procedures are statistically different, even though they share the same spirit of encompassing tests.[1] We would like to point out the difference between the second and third procedures based on equation (4). The probability of not rejecting the null under the second procedure is:

$$pr[\beta_2 \in (-t^*, t^*)], \tag{6}$$

while the same probability under the third procedure is

$$pr[\beta_2 \in (-t^*, t^*) \mid \beta_1 = 1], \tag{7}$$

with t^* being a critical point of a t-distribution multiplied by a standard error estimate of $\hat{\beta}_2$. Note that (6) is in the form of marginal probability; (7) in the form of conditional probability.[2] The conditional event $\beta_1 = 1$ in (7) implies that model 1 receives full weight in predicting y, leaving smaller room for model 2's independent role. As a result, conditional on $\beta_1 = 1$, the event that $\beta_2 \in (-t^*, t^*)$, that is, that model 2 does not play a significant independent role in predicting y is more likely. To sum up, the conditional probability (7) under the third testing procedure of not rejecting the null tends to be higher than the marginal probability (6) under the second procedure. Also, with regard to the role of H_1 and H_2, it is worth noting that within the context of the second test procedure they are treated in a symmetric way, implying that the same form of regression equation (4) can be used in testing both $H_1: \beta_2 = 0$ and $H_2: \beta_1 = 0$. However, within the context of the third test procedure they are treated in an asymmetric way, implying that in testing H_1 against H_2 we use the following regression equation:

$$y_t - f_{2t} = \beta_1 f_{1t} + u_t \tag{8}$$

By interchanging the roles of H_1 and H_2, the forecast-encompassing test of H_2 against H_1 can be carried out in exactly the same way.

[1]Obviously, the first test is different from others in that it is a joint hypothesis testing procedure.

[2]You might wish to understand the probability statement on β from a Bayesian perspective: probability based on Bayesian posterior on β.

III. Implementation of the Wald Test

To begin with, let us summarize each of the three test procedures above. Based on the regression equation:

$$y_t = \beta_1 f_{1t} + \beta_2 f_{2t} + u_t, \qquad t = t_0 + m, \cdots, T,^3 \tag{9}$$

the first test is a joint hypothesis test of

$$H_0\colon \beta_1 = 1 \quad \text{and} \quad \beta_2 = 0.$$

The second test is to test $H_0\colon \beta_2 = 0$ using the same equation (9). The third one is to test $H_0\colon \beta_2 = 0$ based on the regression equation:

$$y_t - f_{1t} = \beta_2 f_{2t} + u_t, \qquad t = t_0 + m, \cdots, T. \tag{10}$$

Under the null hypothesis that model 1 is the true model, but model 2 is not, the error term u_t in equation (9) is equal to that in equation (10) for each $t = t_0 + m, \cdots, T$. The common u_t can be written as

$$u_t = y_t - f_{1t}. \tag{11}$$

When we are making m periods ahead forecasts, then f_{1t} can be written as

$$f_{1t} - x'_{1t}\, \hat{\alpha}^{t-m} \tag{12}$$

where $\hat{\alpha}^{t-m}$ is the least squares estimate of α in equation (2) using observations $t = 1, \cdots, t-m$. Let

$$\underset{(t-m)\times d_1}{X_{1t-m}} - \begin{bmatrix} x'_{11} \\ \vdots \\ x'_{1t-m} \end{bmatrix},$$

[3] The reason why t runs from $t_0 + m$ in equation (9) will become clear in the following section.

$$\underset{(t-m)\times 1}{Y_{t-m}} = \begin{bmatrix} y_1 \\ \vdots \\ y_{t-m} \end{bmatrix},$$

$$\underset{(t-m)\times 1}{U_{1t-m}} = \begin{bmatrix} u_{11} \\ \vdots \\ u_{1t-m} \end{bmatrix}. \tag{13}$$

Then

$$\hat{\alpha}^{t-m} = (X'_{1t-m} X_{1t-m})^{-1} X'_{1t-m} Y_{t-m}$$

$$= \alpha + (X'_{1t-m} X_{1t-m})^{-1} X'_{1t-m} U_{1t-m}.$$

Therefore, under the null hypothesis, u_t can be written as

$$u_t = y_t - f_{1t}$$

$$= x'_{1t}\alpha + u_{1t} - x'_{1t}[\alpha + (X'_{1t-m} X_{1t-m})^{-1} X'_{1t-m} U_{1t-m}] \tag{14}$$

$$-u_{1t} - x'_{1t}(X'_{1t-m} X_{1t-m})^{-1} X'_{1t-m} U_{1t-m}$$

and

$$Var(u_t) - \sigma_1^2 [1 + x'_{1t}(X'_{1t-m} X_{1t-m})^{-1} X'_{1t}]. \tag{15}$$

We notice that the variance of the "forecast error" u_t is composed of two parts: the variance of the "model error" u_{1t} and the variance reflecting that the true model is estimated. If the value of x_{1t} does not change much as t changes, we can reasonably approximate the variance of u_t as:

$$Var(u_t) \approx \sigma_1^2 \left(1 + \frac{1}{t-m}\right). \tag{16}$$

Note that the variance component resulting from estimation, approximated as $\sigma_1^2/(t-m)$, disappears as t increases, which is in line with the consistency property of least squares estimators.

However, for finite samples, the forecast error u_t shows heteroscedasticity.

For $t>s$, we can also compute the covariance between u_t and u_s:

$$
\begin{aligned}
\mathrm{Cov}(u_t,u_s) &= \sigma_1^2 x'_{1s}(X'_{1t-m}X_{1t-m})^{-1}x_{1t}[1-1_{(t-m>s)}] \\
&= \sigma_1^2 x'_{1s}(X'_{1t-m}X_{1t-m})^{-1}x_{1t}1_{(1<t-s>m)} \\
&= \begin{cases} \sigma_1^2 x'_{1s}(X'_{1t-m}X_{1t-m})^{-1}x_{1t} & \text{if } 1<t-s<m \\ 0 & \text{if } t-s\geq m. \end{cases}
\end{aligned}
\tag{17}
$$

where $1_{(\cdot)}$ is an indicator function taking value 1 if the condition inside the parentheses is met, and 0 otherwise. If $m=1$, that is, if we are making one-period ahead forecasts, then the forecast errors are uncorrelated. If $m>1$, then we have correlation among nearby forecast errors. But even in this case ($m>1$), the correlation is expected to be small. Note that if the value of x_{1t} is relatively stable over time and if t is large, then both the heteroscedasticity and autocorrelation of u_t's are not serious.

In applying the Wald test in the next section, we may consider three types of variance estimates of the least squares estimators of β_1 and β_2 in equations (9) and (10). Write the equations (9) and (10) in a general matrix form

$$Y=X\beta+U, \tag{18}$$

where the errors are heteroscedastic and serially correlated up to order $m-1$ with m the forecast horizon. Note that when we are making one-period ahead forecasts ($m=1$), the errors are only heteroscedastic.

The conventional variance estimator of $\hat{\beta}$, the least squares estimator of β, is

$$\hat{\mathrm{Var}}(\hat{\beta})=(X'X)^{-1}\hat{\sigma}^2 \tag{19}$$

where

$$\hat{\sigma}^2=\frac{\hat{U}'\hat{U}}{T-K}$$

with $\hat{U}=(\hat{u}_1,\cdots,\hat{u}_T)'$ the least squares residual vector; K the number of the columns in X. The White's (1980) heteroscedasticity consistent variance estimator of $\hat{\beta}$ is

$$\hat{\mathrm{Var}}(\hat{\beta})=(X'X)^{-1}[\sum_{t=1}^{T}\hat{u}_t^2 x_t x'_t](X'X)^{-1} \tag{20}$$

where $\hat{u}_t$ is the t^{th} least squares residual and x'_t is the $1\times K$ row vector corresponding to the t^{th} observation.

The Newey and West (1987) variance estimator of β is

$$\hat{\mathrm{Var}}(\hat{\beta})=(X'X)^{-1}[\sum_{t=1}^{T}u_t^2 x_t x'_t+w_1\sum_{t=2}^{T}\hat{u}_t\hat{u}_{t-1}(x_t x'_{t-1}+x_{t-1}x'_t)$$

$$+\cdots+w_{m-1}\sum_{t=m}^{T}\hat{u}_t\hat{u}_{t-m+1}(x_t x'_{t-m+1}+x_{t-m+1}x'_t)](X'X)^{-1} \tag{21}$$

where $w_j=1-(j/m)$, $j=1,\cdots,m-1$. In the above formula, the term beginning with w_j within the square bracket is introduced to capture the j^{th} serial correlation of the error terms, $j=1,\cdots,m-1$. By the form of the weight function w_j, we readily notice that higher order serial correlations are receiving less and less attention. When $m=1$, the additional terms beyond $\sum_{t=1}^{T}\hat{u}_t^2 x_t x'_t$ drop and the estimator reduces to the previous White heteroscedasticity consistent variance estimator. If we treat each serial correlation symmetrically, we would like to make $w_j=1$ for all $j=1,\cdots,m-1$. The resulting estimator was used by Fair and Shiller (1990).

The null hypotheses of interest can be written in a general form as

$$H_0\colon A\beta=0,$$

where matrix A has full row rank of q. Note that in the first test, $q=2$, and in the second and the third tests, $q=1$. Under the null hypothesis, the Wald test statistic

$$W=\hat{\beta}'A'[A\hat{\mathrm{Var}}(\hat{\beta})A']^{-1}A\hat{\beta} \tag{22}$$

has an asymptotic Chi-square distribution with q degrees of freedom if the variance estimator of $\hat{\beta}$ is consistent.

IV. Monte-Carlo Simulations

To compare the performance of three different tests in finite samples, we conducted a small simulation study.

Data on y are generated according to the following scheme:

$$y_t = \theta_0 + \theta_1 x_{1t} + \theta_2 x_{2t} + \theta_3 x_{12t} + u_t, \quad t = 1, \cdots, T \tag{23}$$

where x_j is uniformly distributed over $[0, l_j]$, $j = 1, 2, 12$ with l_j chosen such that $\mathrm{Var}(x_j) = \sigma_j^2$, that is, $l_j = 2/3\,\sigma_j$; u_t is normally distributed with mean 0 and variance σ^2; X_1, X_2, X_{12} and U are mutually independent.

Two non-nested models considered are

$$H_1\colon\ y_t = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{12t} + u_{1t} \tag{24}$$

and

$$H_2\colon\ y_t = \gamma_0 + \gamma_1 x_{2t} + \gamma_2 x_{12t} + u_{2t} \tag{25}$$

By using observations $t = 1, \cdots, R$, we can estimate $\hat{\alpha}_0^R, \hat{\alpha}_1^R, \hat{\alpha}_2^R$ by running OLS to model H_1; $\hat{\gamma}_0^R, \hat{\gamma}_1^R, \hat{\gamma}_2^R$ by running OLS to model H_2. Under each model, the m period ahead forecast on y at time R can be obtained as follows (see Liang and Ryu (1996)):

$$H_1\colon\ f_{1R-m} = \hat{\alpha}_0^R + \hat{\alpha}_1^R x_{1R+m} + \hat{\alpha}_2 x_{12R-m}, \tag{26}$$

and

$$H_2\colon\ f_{2R+m} = \hat{\gamma}_0^R + \hat{\gamma}_1^R x_{2R-m} + \hat{\gamma}_2^R x_{12R-m}. \tag{27}$$

Let R start from t_0. We readily notice that $t_0 \geq 3$ is needed. We fix $t_0 = 5$ in our simulation study. Also, $R < T - m$ is required.

Now let f_1 and f_2 be $(T - m - t_0 + 1) \times 1$ vectors of (f_{1R+m}), (f_{2R-m}) respectively, each element of which is recursively obtained as R runs from $R = t_0$ to $R = T - m$. Let Y_p be the corresponding $(T - m - t_0 + 1) \times 1$ sub-vector of y. In terms of the combining regression

$$Y_p = \beta_1 f_1 + \beta_2 f_2 + U \tag{28}$$

the null hypotheses that H_1 is the true model takes the following three different forms under each of the three testing procedures:

$$H^1_0: \beta_1=1, \quad \beta_2=0;$$

$$H^2_0: \beta_2=0;$$

$$H^3_0: \beta_2=0 \quad \text{conditional on} \quad \beta_1=1.$$

In our simulation study, we consider three parameter settings:

$$\text{setting } A: \quad \sigma_1=1, \quad \sigma_2=0;$$

$$\text{setting } B: \quad \sigma_1=1, \quad \sigma_2=1;$$

$$\text{setting } C: \quad \sigma_1=1/2, \quad \sigma_2=1/2.$$

Other parameters are common to all three settings: $\theta_0=\theta_1=\theta_2=\theta_3=1$, $\sigma_{12}=1$, $\sigma=1$.

Under setting A, model H_1 is in fact the true model, while under B and C, model H_1 is not true and model H_2 also contains a useful piece of information which is not contained in H_1. Under setting B, model H_1 deviates more from the true model than under setting C.

Sample sizes $T-25$, 50, 100, 200, 500, (1,000 for setting A) forecast horizons $m=1$, 4 and significance levels $\alpha=5\%$, 10% are considered. Random numbers are generated through RNDN for $N(0,1)$ and RNDU for uniform variables using GAUSS. Results are in Tables 1 through 6. These tables show the number of rejections out of 1,000 replications. For $m=1$, we used two types of covariance estimates: the conventional one (C) together with White's heteroscedasticity consistent one (W). Note that when $m-1$, we do not have any serial correlation among forecast errors under the null hypothesis. For $m-4$, we used two additional covariance estimates as used in fair and Shiller (FS), and Newey-West (NW). Under setting A, table entries are related to empirical sizes of tests, while under setting B and C, table entries are related to empirical power.

As the sample size T increases, the empirical size approaches the hypothesized nominal size α. But the differences between the empirical size and nominal size are quite huge even for $T-200$. In general, the differences are larger for $m=4$ than for $m=1$, and also

TABLE 1

Setting A, $m=1$

		Test 1		Test 2		Test 3	
		C	W	C	W	C	W
$T-25$	$\alpha-5\%$	98	168	82	128	73	84
	10%	145	237	134	187	118	129
$T=50$	$\alpha=5\%$	86	91	87	87	50	57
	10%	145	165	147	157	114	112
$T=100$	$\alpha=5\%$	82	83	82	69	58	63
	10%	148	144	133	123	116	121
$T-200$	$\alpha-5\%$	80	58	74	63	45	43
	10%	123	113	133	124	101	98
$T=500$	$\alpha=5\%$	73	64	66	59	66	62
	10%	136	126	127	120	121	122
$T=1{,}000$	$\alpha=5\%$	59	56	61	51	46	44
	10%	103	89	101	91	92	93

TABLE 2

Setting A, $m=4$

		Test 1				Test 2				Test 3			
		C	W	FS	NW	C	W	FS	NW	C	W	FS	NW
$T=25$	$\alpha=5\%$	180	280	339	371	147	209	263	252	100	113	178	152
	10%	263	345	380	437	221	301	317	315	165	295	240	221
$T-50$	$\alpha-5\%$	123	143	225	159	98	113	137	122	83	81	113	97
	10%	192	202	290	237	177	183	216	192	151	147	182	150
$T=100$	$\alpha=5\%$	104	90	138	104	69	69	77	69	71	69	71	71
	10%	153	155	195	147	133	123	129	125	117	115	133	121
$T=200$	$\alpha=5\%$	87	83	96	76	101	91	84	85	72	66	60	69
	10%	164	154	155	146	172	150	169	155	128	132	131	133
$T-500$	$\alpha-5\%$	77	67	57	63	76	67	64	69	57	56	58	56
	10%	127	124	124	114	126	110	115	108	126	126	115	127
$T=1{,}000$	$\alpha=5\%$	79	72	66	69	56	58	50	55	56	54	54	55
	10%	118	120	114	113	110	112	106	112	116	110	116	114

TABLE 3

Setting B, $m=1$

		Test 1		Test 2		Test 3	
		C	W	C	W	C	W
$T-25$	$\alpha-5\%$	637	742	700	762	82	93
	10%	720	800	765	814	141	167
$T=50$	$\alpha=5\%$	883	915	916	932	90	105
	10%	916	935	948	954	151	163
$T=100$	$\alpha=5\%$	991	991	995	994	149	156
	10%	995	993	998	995	231	242
$T-200$	$\alpha-5\%$	1,000	999	1,000	1,000	294	309
	10%	1,000	1,000	1,000	1,000	418	433
$T=500$	$\alpha=5\%$	1,000	1,000	1,000	1,000	690	707
	10%	1,000	1,000	1,000	1,000	792	799
$T=1{,}000$	$\alpha=5\%$	1,000	1,000	1,000	1,000	913	915
	10%	1,000	1,000	1,000	1,000	953	956

TABLE 4

Setting B, $m=4$

		Test 1				Test 2				Test 3			
		C	W	FS	NW	C	W	FS	NW	C	W	FS	NW
$T=25$	$\alpha=5\%$	595	722	549	775	632	727	670	753	120	136	198	185
	10%	671	786	569	823	727	795	718	801	193	212	277	256
$T=50$	$\alpha=5\%$	881	920	881	933	903	926	911	922	158	163	172	162
	10%	917	943	893	940	934	940	935	945	232	253	237	226
$T=100$	$\alpha=5\%$	994	993	986	990	999	994	993	995	173	176	181	177
	10%	998	994	997	997	1,000	997	995	997	247	259	271	256
$T-200$	$\alpha-5\%$	1,000	1,000	999	999	1,000	1,000	1,000	1,000	329	337	331	329
	10%	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	431	449	435	433
$T=500$	$\alpha=5\%$	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	671	680	667	674
	10%	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	784	795	778	789

TABLE 5

Setting C, $m=1$

		Test 1		Test 2		Test 3	
		C	*W*	*C*	*W*	*C*	*W*
$T=25$	$\alpha=5\%$	398	489	419	502	71	88
	10%	486	587	523	589	116	141
$T=50$	$\alpha=5\%$	591	632	665	674	67	70
	10%	689	717	752	768	108	114
$T=100$	$\alpha=5\%$	852	855	891	893	70	66
	10%	903	907	938	937	132	135
$T=200$	$\alpha=5\%$	982	978	988	985	100	106
	10%	990	987	992	992	160	160
$T=500$	$\alpha=5\%$	1,000	999	1,000	1,000	186	190
	10%	1,000	1,000	1,000	1,000	280	282

TABLE 6

Setting C, $m=4$

		Test 1				Test 2				Test 3			
		C	*W*	*FS*	*NW*	*C*	*W*	*FS*	*NW*	*C*	*W*	*FS*	*NW*
$T=25$	$\alpha=5\%$	400	514	493	615	407	496	518	539	103	123	185	175
	10%	495	591	531	681	502	571	592	615	173	199	243	223
$T=50$	$\alpha=5\%$	575	634	684	693	651	692	704	683	96	101	141	115
	10%	685	735	740	754	732	754	770	775	167	169	180	167
$T=100$	$\alpha=5\%$	807	847	836	847	868	861	873	867	101	100	97	97
	10%	891	903	887	893	899	911	905	907	158	172	148	154
$T=200$	$\alpha=5\%$	983	975	973	972	996	989	988	989	93	87	89	85
	10%	994	991	986	987	999	998	993	997	143	145	141	133
$T=500$	$\alpha=5\%$	1,000	998	998	997	1,000	998	996	999	163	167	165	167
	10%	1,000	1,000	999	998	1,000	1,000	999	999	261	263	261	265

larger for tests 1, 2 than for test 3. In Table 1, using White's heteroscedasticity consistent covariance estimator (*W*) does not always improve the finite sample approximation using the conventional formula (*C*). In fact, until the sample size reaches $T=$

100, W has shown even poorer performances. But as we expect, as the sample size increases, W offers a better approximation than C.

In Table 2, neither FS nor NW offers a better approximation than W. In fact, both FS and NW have shown even poorer performance until the sample size reaches $T-200$. Even for $T-500$, 1,000, we could not see any significant improvement of FS or NW over W. Regarding the relationship between C and W, the same pattern is observed as in Table 1: W is worse than C until the sample size reaches $T=100$. After that, W outperforms C. In Tables 3 and 5, using W gives higher power than using C. But this may simply reflect that the size of the test using W is inflated in small sample sizes (see Table 1). In Tables 4 and 5 note that using NW offers much higher power than using FS. Even though sizes of tests using FS and NW have both inflated by about the same rate, NW offers much higher power.

Also as T increases, empirical power increases for all three tests. But the empirical power of test 3 is extremely lower than that of tests 1 and 2, which is consistent with our earlier explanations. Considering size and power together, we would choose tests 1, 2 over test 3.

V. Applications to the Choice of Consumption Function

Gaver and Geisel (1974) proposes two forms of a consumption function

$$H_1: C_t = \beta_1 + \beta_2 Y_t + \beta_3 C_{t-1} + U_{1t}$$

and

$$H_2: C_t - \gamma_1 + \gamma_2 Y_t + \gamma_3 Y_{t-1} + U_{2t},$$

where C_t is the real aggregate consumption in year t, and Y_t is the real aggregate income in the same year. Model H_1 implies that the effects of changes in income on consumption persist for many years, while model H_2 implies that consumption responds to changes in income only over the recent two years.

Using aggregate annual U.S. data for 1929 through 1989, we performed each of the three proposed forecast encompassing tests.

TABLE 7

	Test 1				Test 2				Test 3			
	C	*W*	*FS*	*NW*	*C*	*W*	*FS*	*NW*	*C*	*W*	*FS*	*NW*
H_1 against H_2												
$m-1$	4.7*	4.2	4.2	4.2	1.4	1.6	1.6	1.6	3.3*	2.4	2.4	2.4
$m-4$	3.2	4.1	1.4	1.9	0.9	1.1	0.4	0.5	2.3	2.2	0.6	0.8
H_2 against H_1												
$m=1$	31.2**	26.3**	26.3**	26.3**	6.1**	6.7**	6.7**	6.7**	23.3**	21.5**	21.5**	21.5**
$m=4$	36.1**	50.6**	17.8**	24.1**	16.4**	22.0**	6.8**	9.9**	15.5**	20.7**	4.7**	7.0**

Notes: (1) All three tests are to be evaluated according to chi-square distribution.
(2) The degrees of freedom are 2, 1, and 1 for Test 1, 2, 3, respectively.
(3) C—using conventional variance formula
W=using White's heteroscedasticity consistent variance formula
FS=using heteroscedasticity and autocorrelation consistent variance formula as in Farr and Shiller
NW=using heteroscedasticity and autocorrelation consistent variance formula as in Newey and West
(4) $m-1$: 1-period ahead forecast
$m=4$: 4-period ahead forecast
(5) H_1: $C_t = \beta_0 + \beta_1 Y_t + \beta_2 C_{t-1} + U_{1t}$
H_2: $C_t = \gamma_0 + \gamma_1 Y_t + \gamma_2 Y_{t-1} + U_{2t}$
(6) *: significant at $\alpha = 10\%$
**: significant at $\alpha - 5\%$.

For each test, we used four different variance estimators: conventional, White's, Fair and Shiller's, and Newey and West's. For comparison purposes, we also carried out Davidson and MacKinnon's J-test. The encompassing test results are summarized in Table 7.

Our proposed encompassing tests cannot reject the null model of H_1 at $\alpha=5\%$ significance level. Even at $\alpha=10\%$ significance level, we cannot reject the null model of H_1 except for two cases. On the other hand, we always reject the null model of H_2 at $\alpha=5\%$ significance level. Based on our encompassing test results, we clearly select model H_1 over H_2 and conclude that in the U.S. changes in income affect consumption over many periods.

TABLE 8
OUTCOMES AND CONCLUSIONS OF THE PAIRWISE FORECAST-ENCOMPASSING TEST

Hypothesis H_2	Hypothesis H_1	
	Do not reject H_1	Reject H_1
Do not reject H_2	(i) Reject neither H_1 nor H_2	(iii) Do not reject H_2, but reject H_1
Reject H_2	(ii) Do not reject H_1, but reject H_2	(iv) Reject both H_1 and H_2

The results of Davidson and MacKinnon's J-test are as follows:

$$H_1 \text{ against } H_2\text{: } t\text{-value}=8.3^{**}$$

$$H_2 \text{ against } H_1\text{: } t\text{-value}-11.4^{**}$$

The J-test rejects both null models, and thus is not able to suggest a single model.

VI. Implications of the Forecast-Encompassing Test Results

The above analysis can be regarded as applying to a situation where economists are seeking to discriminate between two competing forecasting formulas such as (2) and (3) in section II. It investigates if there is a significant contribution of f_2 to predictive power after controlling for f_1. In this sense, H_1 or f_1 is tested against H_2 or f_2. The problem is analogous to comparing two non-nested specifications in econometrics (McAleer 1987), and the choice of alternative hypothesis plays a critical role because it affects the power of the test when the null is not valid (Chong and Hendry 1986). When the roles of H_1 and H are symmetrically considered, four outcomes are possible (see Table 8): (i) reject neither H_1 nor H_2; (ii) do not reject H_1, but reject H_2; (iii) do not reject H_2, but reject H_1; and (iv) reject both H_1 and H_2. Accordingly, it is possible to accept both forecasting formulas or to reject both. Only in cases (ii) and (iii), it is possible to discriminate between the two rival formulas.

The four possible outcomes we have mentioned lead to some rather valuable insights on cross-model evaluations in econometrics and modelling strategy in composite forecasts. From an econometric perspective, a failure to achieve encompassing acknowledges the possibility of an incomplete model (Hendry 1983; and Hendry and Richard 1983), thereby time is better spent improving the model specifications. Conversely, an ability to achieve encompassing recognizes that the encompassing model is satisfactory at least at the current stage. According to Table 8, when model 1 serves as the mull and model 2 as the alternative, outcomes (iii) and (iv) acknowledge the possibility that model 1 is misspecified, implying that a better model can be constructed (Chong and Hendry 1986). On the other hand, outcome (ii) admits that model 1 stands up to the forecast-encompassing criterion. Outcome (i) suggests that we do not yet have enough data to reject either model with confidence.

In the case of composite forecasts, a failure to achieve encompassing suggests that the strategy of combining forecasts is justified (Hallman and Kamstra 1989; and Fair and Shiller 1990). On the other hand, an ability to achieve encompassing offers a logical reason that researchers may concentrate on the forecast which dominates others (Granger 1989). In this sense, an ability or a failure to achieve encompassing provides researchers a useful guide to resolve an important but generally neglected issue in the literature of composite forecasts: that is, under what conditions is combining most useful (Armstrong 1989)? Clearly, if a model encompasses others in forecast, then using a combination of forecasts will not achieve significant information gain. In this case, a single best model alone is capable of forecasting the variable of interest "significantly well," and therefore it is unnecessary to consider other models. On the other hand, should we end up rejecting both H_1 and H_2 (outcome (iv)), it is implied that a composite forecast containing both f_1 and f_2 can outperform each of the two individual forecasts. In this situation, the sample is not entirely consistent with H_1 or H_2 individually. Therefore, a pragmatic attempt would be to use the artificial combining regression (4) to obtain a better description of the data than does either individual forecast. Fisher and McAleer (1979) provides an analogous argument on this aspect in a non-nested hypothesis testing context. Liang and Ryu (1996) provides a constructive way of dealing with the combining regression like equation (4).

In sum, outcome (iv) suggests that both model 1 and model 2 forecasts contain useful information that are capable of generating a composite forecast with smaller forecasting error than does either f_1 or f_2. Outcome (ii) and (iii) acknowledge that using f_1 or f_2 alone is sufficient to generate forecasts of Y with "almost" the same degree of accuracy as both f_1 and f_2 combined, implying that it is not worthwhile pursuing a combining strategy in either case. Finally, outcome (i) suggests that there is insufficient information to discriminate between the two forecasts. When f_1 and f_2 are highly colinear, β_1 and β_2 in (4) may not be separately identified (Kennedy 1989; and Fair and Shiller 1990). This coincides with the case of unstable combining weights in the literature of composite forecasts (*e.g.*, Kang (1986)).

VII. Concluding Remarks

The development of the method of the combining regression and its application to economics has led to comparisons among the forecasting abilities of rival models. Using the forecast-encompassing principle, it is possible to test whether one specification dominates others. On the other hand, a great number of theoretical and empirical works in forecasting have shown and demonstrated the superiority of the composite forecasts over individual forecasts under some criteria.

This paper examines the role of forecast-encompassing principles in model-specification searches through the use of linear composite forecasts. In this way the approach has similarities to non−nested hypotheses testing in econometrics. Based on the outcomes and conclusions of the pairwise forecast-encompassing tests, this paper also outlines a conceptual framework to provide some useful insights on cross-model evaluations in econometrics and the selection of predictors in composite forecasts. The conventional wisdom focuses on the former aspect but not the latter. Once the two-way interaction is established, the complementary role of composite forecasts and the forecast-encompassing principle can be clearly presented.

Overall, the contribution of this paper is twofold. First, it clarifies the complementary role of the forecast-encompassing principle and composite forecasts. Second, it studies the relationship of forecast-encompassing principles to composite forecasts and provides three different ways of performing the encompassing test. Test outcomes guide researchers to choose component forecasts and thus to avoid blind pooling in the combining regression. Given the wide advocacy of the combining forecasts and the advent of inexpensive forecasting software for personal computers, clearly this is an important issue; it is, however, an issue that has been rather neglected.

(*Received 19 February 2003; Revised 7 April 2004*)

References

Andrews, D. W. K. "Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica* 59 (No. 3 1991): 817-58.

Armstrong, J. S. "Combining Forecasts: The End of the Beginning of the Beginning of the End." *International Journal of Forecasting* 5 (1989): 585-8.

Bates, J. M., and Granger, C. W. J. "The Combination of Forecasts." *Operational Research Quarterly* 20 (1969): 451-68.

Bunn, D. W. "Forecasting with More than One Model." *Journal of Forecasting* 8 (1989): 161-6.

Charemza, W. W. "Large Econometric Models of an Eastern European Economy, A Critique of the Methodology." *Economic Modelling* 8 (1991): 45-62.

Chong, Y. Y., and Hendry, D. F. "Econometric Evaluation of Linear Macroeconometric Models." *Review of Economic Studies* 53 (No. 4 1986): 671-90.

Clemen, R. T. "Combining Forecasts: A Review and Annotated Bibliography." *International Journal of Forecasting* 5 (1989): 559-83.

Cooper, J. P., and Nelson, C. R. "The Ex-Ante Prediction Performance of the St. Louis and FRB-MIT-PENN Econometric Models, and Some Results on Composite Predictions." *Journal of Money, Credit, and Banking* 7 (No. 1 1975): 1-32.

Davidson, R., and MacKinnon, J. "Several Tests for Model Specification in the Presence of Alternative Hypotheses."

Econometrica 49 (No. 3 1981): 781-93.

Diebold, F. X. "Forecast Combination and Encompassing: Reconciling Two Divergent Literatures." *International Journal of Forecasting* 5 (1989): 589-92.

Fair, R. C., and Shiller, R. J. "Comparing Information in Forecasts from Econometric Models." *American Economic Review* 80 (No. 3 1990): 375-89.

Fisher, G., and McAleer, M. "On the Interpretation of the Cox Test in Econometrics." *Economics Letters* 4 (1979): 145-50.

Fisher, P. G., and Wallis, K. F. "The Historical Tracking Performance of U.K. Macroeconometric Models, 1978-85." *Economic Modelling* 8 (1990): 45-62.

Gaver, K., and Geisel, M. "Discriminating Among Alternative Models: Bayesian and Non-Bayesian Methods." In P. Zarembka (ed.), *Frontiers in Econometrics*. NY: Academic Press, 1974.

Granger, C. W. J. "Combining Forecasts-Twenty Years Later." *Journal of Forecasting* 8 (1989): 167-73.

Granger, C. W. J., and Ramanathan, R. "Improved Methods of Forecasting." *Journal of Forecasting* 3 (1984): 197-204.

Hallman, J., and Kamstra, M. "Combining Algorithms Based on Robust Estimation Techniques and Co-Integrating Restrictions." *Journal of Forecasting* 8 (1989): 189-98.

Hendry, D. F. "Econometric Modelling: The Consumption Function in Retrospect." *Scottish Journal of Political Economy* 30 (No. 3 1983): 193-200.

Hendry, D. F., and Richard, J. F. "The Econometric Analysis of Economic Time Series (with Discussions)." *International Statistical Review* 51 (1983): 111-63.

Kang, H. "Unstable Weights in the Combination of Forecasts." *Management Science* 32 (1986): 683-95.

Kennedy, P. "Non-Nested Hypothesis Tests: A Diagrammatic Exposition." *Australian Economic Papers* 28 (No. 52 1989): 160-5.

Liang, K. Y., and Ryu, K. "Selecting the Form of Combining Regressions Based on Recursive Prediction Criteria." In J. C. Lee, W. O. Johnson, and A. Zellner (eds.), *Modelling and Prediction Honoring Seymour Geisser*. Spriner-Verlag, 1996.

McAleer, M. H. "Specification Tests for Separate Models: A Survey." In M. L. King, and D. E. A. Goles (eds.), *Specification*

Analysis in the Linear Model. London: Routledge & Kegan, pp. 146-95, 1987.

Mizon, G. E., and Richard, J. F. "The Encompassing Principle and Its Application of Non-Nested Hypotheses." *Econometrica* 54 (No. 3 1986): 657-78.

Nelson, C. R. "The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy." *American Economic Review* 62 (No. 5 1972): 902-17.

Newbold, P., and Granger, C. W. J. "Experience with Forecasting Univariate Time Series and the Combination of Forecasts (with Discussions)." *Journal of the Royal Statistical Society, Series A* 137 (1974): 131-49.

Newey, W. K., and West, K. D. "A Simple Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55 (No. 3 1987): 703-8.

Wallis, K. F. "Macroeconomic Forecasts: A Philosophical Basis and Some Current Issues." *International Journal of Forecasting* 5 (1989): 605-9.

White, H. "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity." *Econometrica* 48 (No. 3 1980): 817-38.

Winkler, R. L. "Combining Forecasts: A Philosophical Basis and Some Current Issues." *International Journal of Forecasting* 5 (1989): 605-9.